



Interactions audio-tactiles et perception de la parole : Comparaisons entre sujets aveugles et voyants

Christian Cavé, Marc Sato, Lucie Ménard, Annie Brasseur

► To cite this version:

Christian Cavé, Marc Sato, Lucie Ménard, Annie Brasseur. Interactions audio-tactiles et perception de la parole : Comparaisons entre sujets aveugles et voyants. JEP 2010 - 28e Journées d'Etudes sur la Parole, May 2010, Mons, Belgique. pp.21-24. hal-00541981

HAL Id: hal-00541981

<https://hal.science/hal-00541981>

Submitted on 1 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interactions audio-tactiles et perception de la parole :

Comparaisons entre sujets aveugles et voyants

Christian Cavé¹, Marc Sato², Lucie Ménard³, Annie Brasseur³

¹Laboratoire Parole & Langage, CNRS & Aix-Marseille Université, France

²GIPSA-Lab, Département Parole & Cognition, CNRS & Grenoble Universités

³Département de Linguistique, Université du Québec à Montréal, Canada

Email : christian.cave@lpl-aix.fr

ABSTRACT

The present study investigated whether tactile information obtained manually by touching the speaker's face modulates the decoding of speech. Audio-tactile perception was compared to audio-only perception, and a group of congenitally blind adults was compared to a group of sighted adults. Participants performed a phonemic decision task in three conditions: audio-only, congruent audio-tactile, and incongruent audio-tactile. For the auditory modality, the phoneme sequences were presented either with a background white noise or without noise. The results showed that tactile information relevant to recovering speech gestures improved auditory speech perception in cases of degraded acoustic information. Moreover, the same audio-tactile interactions were found for the blind and sighted listeners, despite possible differences in these groups' sensory skills.

Keywords: speech perception; multimodality; audio-tactile interactions; blindness.

1. INTRODUCTION

On sait depuis longtemps que des informations visuelles modifient le traitement de la parole par exemple en améliorant l'intelligibilité de parole présentée dans le bruit si le visage du locuteur est visible [1-2]. L'effet McGurk [3] est une autre manifestation de l'influence d'informations visuelles sur le décodage de la parole qui a été étudiée pour de nombreuses langues et divers types de sujets, normo entendants, malentendants, avec troubles de parole, etc. (voir [4] pour une synthèse).

On sait aussi depuis longtemps, par la méthode Tadoma [5] utilisée par des sujets sourds et aveugles, que des informations tactiles (perception haptique), obtenues en plaçant une main sur le visage du locuteur, permettent de récupérer les gestes de production de la parole (voir la partie « Méthode » pour détails). Des utilisateurs entraînés peuvent ainsi arriver à un niveau de communication quasi-normal.

Quelques études ont montré que des interactions audio-tactiles pour la parole existaient chez les sujets normaux non entraînés. Ainsi Fowler et Dekle [6] ont présenté à des sujets naïfs non entraînés des syllabes acoustiques couplées avec le contact de la main sur le visage du locuteur. Ils mettent ainsi en évidence une forte

interaction entre les modalités auditive et tactile : l'information tactile influence le décodage de la syllabe auditive et réciproquement, la syllabe auditive le décodage de la syllabe perçue tactilement. De plus, la présentation de paires de syllabes auditive-tactile non cohérentes, produit chez certains sujets un percept illusoire (fusion ou combinaison) de type McGurk. Il faut toutefois noter qu'il y a de fortes différences interindividuelles et que la plupart des sujets ne perçoivent pas de réponses illusoires. Plus récemment, Gick et collègues [7] ont montré que l'information tactile manuelle améliorerait la perception de la parole tant visuelle qu'auditive chez des sujets non entraînés. Dans leur étude, les participants devaient percevoir des disyllabes présentés visuellement ou auditivement dans un bruit blanc continu, avec la main placée sur le visage du locuteur comme dans la méthode Tadoma. La comparaison entre la condition bimodale et les conditions unimodales montrent que l'information tactile améliore la perception de la parole d'environ 10%. Pris ensemble, ces résultats soulèvent d'importantes questions sur les relations entre la modalité auditive et les autres modalités sensorielles et sur un possible couplage fonctionnel entre systèmes de perception et de production de la parole [8-10].

Pour cette étude, nous avons utilisé la méthode Tadoma pour évaluer l'interaction entre information tactile et information auditive –cohérente ou non cohérente– lors de la perception de la parole en comparant une situation audio-tactile avec une situation audio seul ou tactile seul. Pour cela, il a été demandé aux sujets de réaliser une tâche d'identification phonémique dans trois conditions de présentation: 'audio seul', 'audio-tactile cohérent' ou 'audio-tactile non cohérent'. De plus, une condition 'tactile seul' avait pour but de déterminer dans quelle mesure la seule information tactile permettait d'identifier l'item articulé silencieusement. De façon à évaluer si le niveau de sensibilité tactile peut jouer sur la récupération des gestes de production de la parole et moduler les interactions inter modales, nous avons comparé deux groupes de dix sujets aveugles congénitaux et de dix sujets à vision normale.

Compte tenu des données ci-dessus, nous faisons l'hypothèse qu'une information tactile cohérente devrait améliorer la perception de la parole dans le bruit. Pour la condition audio-tactile, la perception d'un geste articulatoire non cohérent pourrait diminuer

l'intelligibilité de la parole ou même modifier l'expérience auditive du sujet comme lors de l'effet McGurk. La mise en évidence de percepts illusoire de type McGurk lors d'une stimulation audio-tactile non cohérente fournirait de plus un argument en faveur d'une véritable intégration bimodale audio-tactile.

2. MÉTHODE

Participants

Un groupe de dix sujets aveugles congénitaux (moyenne d'âge : 41 ans) et un groupe de dix sujets ayant une vision normale (moyenne d'âge : 27 ans) ont participé à l'étude. Les participants aveugles avaient une cécité de niveau 3, 4 ou 5 selon l'échelle de l'Organisation mondiale de la santé. Tous les participants étaient droitiers, n'avaient pas de troubles d'audition ou de langage connus, avaient le français québécois comme langue maternelle et n'avaient jamais participé à une expérience de même type. Tous ont signé un consentement et ont été payés pour leur participation.

Stimuli

Une locutrice dont la langue maternelle est le français québécois a été enregistrée (44,1 kHz, 16bits) et filmée (30 trames/s ; 720 x 480 pixels) de face sur un fond gris clair. Elle a produit, à plusieurs reprises, les séquences phonémiques /aba/ et /aga/ en maintenant une intensité normale, une intonation neutre et en fermant complètement les lèvres entre chaque réalisation. Une réalisation articulatoire unique de chaque item a été sélectionnée pour produire les 14 stimuli utilisés pour l'expérimentation. Ces stimuli permettent de présenter simultanément la piste audio aux sujets et la piste vidéo à l'expérimentateur (voir Procédure). Pour chaque item, il y avait 7 conditions de présentation :

- tactile seul
- audio seul avec ou sans bruit blanc
- audio-tactile cohérent avec ou sans bruit blanc
- audio-tactile non cohérent avec ou sans bruit blanc

Procédure expérimentale

La procédure expérimentale est inspirée de celle de Fowler et Dekle (1991). Les participants étaient assis face à l'expérimentatrice (AB). Dans la modalité tactile, leur main droite était placée sur son visage de façon à ce que le pouce soit vertical et frôle les lèvres et que les autres doigts soient placés horizontalement sur la mâchoire. Cette position permettait de capter les mouvements des lèvres et de la mandibule lors de la production silencieuse des items /aba/ et /aga/. Pour éviter que les sujets voyants ne regardent l'expérimentatrice, ils portaient un bandeau sur les yeux pendant toute la durée de l'expérience. Pour la modalité auditive, les items /aba/ ou /aga/ étaient présentés par écouteurs mixés ou non avec un bruit blanc à un niveau acoustique confortable.

L'expérimentatrice était assise face au sujet et à un écran d'ordinateur. À chaque essai, l'écran lui indiquait la séquence à prononcer et affichait les mouvements visuels

des articulateurs à produire, ce qui lui permettait d'articuler silencieusement en synchronie avec l'item acoustique présenté au sujet. Avant l'expérimentation, elle était entraînée à articuler silencieusement en synchronie avec les mouvements affichés sur son écran. À aucun moment, pendant le déroulement d'une session, elle n'était au courant de la syllabe présentée de manière acoustique au sujet.

Avant l'expérience, les participants étaient informés qu'il leur serait présenté des items soit auditivement par écouteurs, soit tactilement par contact entre leur main et le visage de l'expérimentatrice, soit par les deux modalités en même temps. Leur tâche était de dire, à chaque essai, s'ils avaient entendu /aba/, /ada/, /aga/ ou n'importe quelle combinaison syllabique (par exemple, /agba/, /abga/ ou /agda/ par exemple). Ils ne recevaient aucune information particulière sur la façon d'interpréter l'information tactile, ni aucune autre consigne. L'expérimentatrice démarrait chaque essai en appuyant sur une touche du clavier tout en ayant la bouche fermée en position neutre. Avant la présentation de l'item auditif au sujet, les instructions relatives à l'articulation à réaliser (i.e., « aba », « aga » ou « ### ») s'affichaient sur son écran pendant 500 ms. L'item acoustique pour le sujet et le visage articulatoire ou non pour l'expérimentatrice étaient alors présentés simultanément. Après chaque présentation, un autre expérimentateur notait la réponse du sujet. De plus, la totalité de la session expérimentale était filmée pour vérification ultérieure si besoin. Chacun des 14 stimuli était présenté 6 fois dans un ordre aléatoire. La procédure expérimentale étant assez contraignante tant pour l'expérimentatrice que pour les participants, ils étaient autorisés à faire de courtes pauses à n'importe quel moment pendant la session expérimentale qui durait environ 20 minutes.

Analyse des résultats

L'ensemble des réponses recueillies a été analysé pour chaque participant et chaque condition. Exceptée pour la condition 'tactile seul', les données ont été traitées par une analyse de variance (Anova) avec pour variable inter-sujets le groupe (voyants, non-voyants) et pour variables intra-sujets la séquence phonémique présentée de manière acoustique (/aba/, /aga/), la production articulatoire silencieuse (/aba/, /aga/, pas de production) et la présence ou non de bruit blanc. Par convention, dans les conditions audio-tactiles non cohérentes, les termes 'réponses correctes' correspondent aux réponses basées sur la modalité auditive. Du fait de l'absence de bruit blanc associé, les données issues de la condition 'tactile seul' ont été traitées séparément des autres conditions par une analyse de variance (ANOVA) avec pour variable inter-sujets le groupe (voyants, non-voyants) et pour variable intra-sujets la syllabe articulée silencieusement (/aba/, /aga/). Pour ces analyses, le niveau de significativité a été fixé à $p < .05$, un test de Mauchly a été effectué de manière à vérifier l'hypothèse de sphéricité des données, enfin des tests de Newman-Keuls ont été utilisés pour les analyses post-hoc.

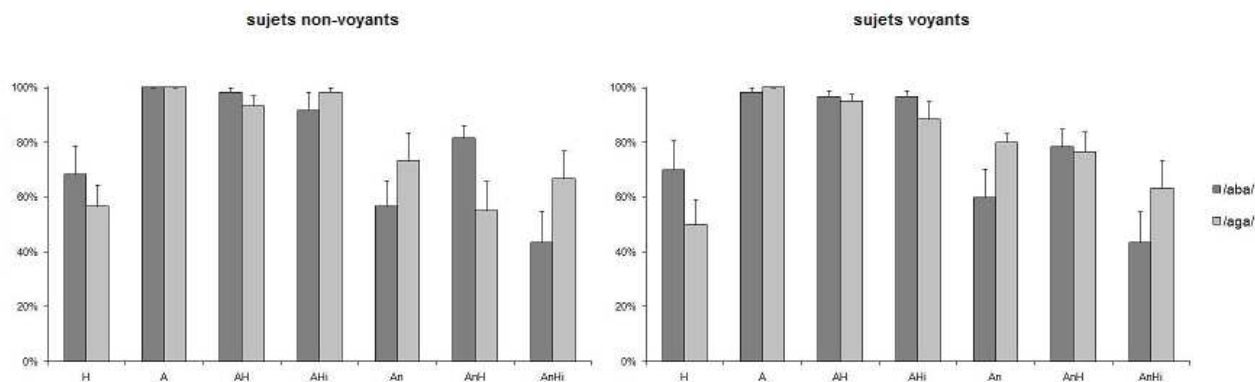


Figure 1. Pourcentage de réponses correctes des participants non-voyants et voyants en fonction des différentes conditions expérimentales (H : tactile seul ; A : audio seul, AH : audio-tactile cohérent, AHi : audio-tactile non cohérent, n : présence de bruit blanc).

3. RÉSULTATS

Le pourcentage de réponses correctes est indiqué dans la Figure 1 pour les deux groupes de sujets.

Condition tactile seul

La syllabe articulée a un effet significatif ($F(1,18) = 5,63$; $p=.03$) sur les réponses. On observe plus de réponses correctes pour /aba/ que pour /aga/ aussi bien pour les aveugles (68% vs 57%) que pour les voyants (70% vs 50%) et il n'y a pas de différence significative entre les deux groupes de sujets ($F(1,18) < 1$).

Condition audio et audio-tactile

Il y a un effet très significatif du bruit ($F(1,18) = 52,07$; $p < .0001$). On observe un plus grand nombre de réponses correctes pour les items présentés sans bruit (96%) que pour les items bruités (65%).

L'item articulé a aussi un effet significatif ($F(2,36) = 5,57$; $p < .008$) ; l'item /aba/ donne lieu à un plus grand nombre de réponses correctes que l'item /aga/ (82% vs 76% ; $p < .02$).

En l'absence de bruit, aussi bien /aba/ que /aga/ sont identifiés de façon quasi parfaite (globalement 99% vs 98%) quelle que soit la condition et il n'y a pas de différence significative entre les conditions.

Par contre, en présence de bruit, les conditions de présentation ont un effet significatif (toutes les comparaisons significatives à au moins $p < .03$) sur le nombre de réponses correctes (voir tableau 1). Ainsi, pour la syllabe /aba/ perçue de manière acoustique en présence de bruit, on observe un nombre de réponses correctes supérieur dans le cas de la condition audio-tactile cohérent par rapport à la condition audio seul (80% vs. 58%, $p < .003$) et la condition audio-tactile non cohérent (80% vs. 43%, $p < .0001$), et dans le cas de la condition audio seul par rapport à la condition audio tactile incohérent (58% vs. 43%, $p < .008$). Pour l'item /aga/, on observe un nombre supérieur de réponses correctes pour les conditions audio tactile cohérent (72% vs. 59%, $p < .03$) et audio seul (77% vs. 59%, $p < .007$) par rapport à la condition audio tactile non cohérent.

.03) et audio seul (77% vs. 59%, $p < .007$) par rapport à la condition audio tactile non cohérent.

Il est important de noter qu'il n'y a pas d'effet principal du groupe ni d'interaction entre la variable groupe et les autres variables.

Tableau 1. Pourcentage de réponses correctes en présence de bruit en fonction des conditions de présentation

	/aba/	/aga/
Audio-tactile cohérent	80%	72%
Audio seul	58%	77%
Audio-tactile non cohérent	43%	59%

4. DISCUSSION

Nos résultats indiquent clairement que l'information tactile provenant des gestes articulatoires telle qu'elle a été traitée par les participants peut moduler la perception de la parole dans des conditions acoustiques dégradées. En effet, on constate que le contact manuel avec le visage de la locutrice associé à une entrée auditive non cohérente diminue l'intelligibilité de la parole non seulement par rapport à la modalité audio seul mais aussi par rapport à la modalité tactile seul, révélant ainsi une certaine interaction entre les informations tactiles et les informations auditives. De même, une augmentation de l'intelligibilité est observée lors de la condition 'audio-tactile cohérent' par rapport aux modalités audio seul et tactile seul.

Toutefois nos résultats ne mettent pas en évidence d'illusions de type McGurk induites par nos stimulus audio-tactiles non cohérents. En effet, le pourcentage de réponses /ada/ ainsi que le pourcentage de réponses de type combinaison n'étaient pas significativement différents entre les conditions audio seul et audio-tactile. Notons cependant que la procédure utilisée ne permet pas

de contrôler l'information tactile perçue par les participants et qu'il en est de même pour les travaux cités [6], [7]. Ce fait pourrait jouer un rôle dans les différences inter- et intra-individuelles constatées. Toutefois, il ne peut en rien expliquer l'augmentation de l'intelligibilité dans la situation audio-tactile cohérent avec bruit.

Nos résultats ne révèlent aucune différence perceptive entre sujets voyants et sujets aveugles. Ce, notamment pour les situations unimodales tactile ou auditive, bien qu'une sensibilité tactile ou auditive supérieure par rapport aux sujets voyants ait été décrite chez des aveugles congénitaux [11-13]. Toutefois, on sait que même des sujets sourds-aveugles entraînés à la méthode Tadoma n'ont pas de meilleurs scores que des sujets standards non entraînés pour l'identification de syllabes sans signification [14, 15]. Il n'est donc pas surprenant de ne pas trouver de différences entre nos deux groupes de sujets pour les situations bimodales, qu'elles soient cohérentes ou non cohérentes.

Plusieurs sujets ont assez rapidement détecté que, dans certains essais, les informations tactiles et auditives n'étaient pas cohérentes. Il semble bien que cela ne se produise pas pour les non cohérences audiovisuelles donnant lieu à l'effet McGurk. Cela pourrait indiquer que les phénomènes d'intégration entre modalité tactile et modalité auditive ne sont pas de même nature ou du moins n'ont pas la même force que ceux intervenant entre l'audition et la vision.

En dépit de l'absence globale d'effet de type McGurk pour la condition audio-tactile non cohérente, certains sujets ont donné des réponses de type fusion /ada/ et de type combinaison telles que /abda/, /abga/, /adga/, ce qui est classique, ou même /adada/, /ababa/, /ania/ et /abia/. Alors que dans notre étude, plusieurs sujets n'ont donné aucune réponse de type fusion ou combinaison, le sujet le plus productif en a donné 32%. Ces observations vont dans le sens de celles de Fowler et Dekle [6] qui rapportent l'existence de percepts illusoire de type McGurk pour des stimulations audio-tactiles non cohérentes, en précisant toutefois qu'il existe de très fortes différences interindividuelles.

5. CONCLUSION

En conclusion, ces résultats démontrent : 1) que lors du processus de traitement de la parole, une information tactile sur les mouvements des articulateurs peut influencer le décodage de l'information auditive présentée simultanément et 2) que les interactions audio-tactiles ne sont pas différentes chez des participants non entraînés en dépit des possibles différences de capacités sensorielles entre sujets voyants et sujets aveugles.

BIBLIOGRAPHIE

[1] Sumby, W.H. and Pollack, I. Visual contribution to speech intelligibility in noise. *Journal of Acoustical Society of America*, 26: 212-215, 1954.

[2] MacLeod, A. and Summerfield, Q. Quantifying the

contribution of vision to speech perception in noise. *British Journal of Audiology*, 21: 131-141, 1987.

[3] McGurk, H. and MacDonald, J. Hearing lips and seeing voices. *Nature*, 264: 746-748, 1976.

[4] Colin, C. and Radeau, M. Les illusions McGurk dans la parole : 25 ans de recherches. *L'Année Psychologique*, 103(3) : 497-542, 2003.

[5] Alcorn, S. The Tadoma method. *Volta Rev.*, 34: 195-198, 1932.

[6] Fowler, C. and Dekle, D. Listening with eye and hand: Crossmodal contributions to speech perception. *J. Exp. Psychol. Hum. Percept. Perform.*, 17: 816-828, 1991.

[7] Gick, B., Jóhannsdóttir, K.M., Gibrael, D. and Mühlbauer, M. Tactile enhancement of auditory and visual speech perception in untrained perceivers, *Journal of Acoustical Society of America*, 123: 72-76, 2008.

[8] Schwartz, J.-L., Sato, M. and Fadiga, L. The common language of speech perception and action: a neurocognitive perspective. *Revue Française de Linguistique Appliquée*, 13(2): 9-22, 2008.

[9] Schwartz, J.-L., Ménard, L., Basirat, A. and Sato, M. The Perception for Action Control Theory (PACT): a perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, sous presse.

[10] Sato, M., Tremblay, P. and Gracco, V. A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, 111(1): 1-7, 2009.

[11] Röder, B. and Neville, H.J. Developmental functional plasticity. In Grafman J, Robertson I (Eds), *Plasticity and rehabilitation. Handbook of Neuropsychology*, vol 9. Elsevier, Amsterdam, pp. 231-270, 2003.

[12] Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R. J., and Belin, P. Pitch discrimination in the early blind. *Nature*, 430: 309, 2004.

[13] Ménard, L., Dupont, S. Baum, S.R. and Aubin, J. Production and perception of French vowels by congenitally blind adults and sighted adults. *Journal of the Acoustical Society of America*, 126: 1406-1414, 2009.

[14] Norton, S. J., Schultz, M. C., Reed, C. M., Braida, L. D., Durlach, N. I., Rabinowitz, W. M. & Chomsky, C. (1977). Analytic study of the Tadoma method: Background and preliminary results. *J. Speech Hear. Res.*, 20: 574-595.

[15] Reed, C. M., Rubin, S. I., Braida, L. D. & Durlach, N. I. (1978). Analytic study of the Tadoma method: discrimination ability of untrained observers. *J Speech Hear Res*, 21: 625-637.